Machine Learning-Based Knee Cartilage Segmentation for Improved Management of Osteoarthritis (OA) with Point of Care Ultrasound (POCUS)

The University of British Columbia

ELEC 523 - Medical Imaging Class Project Report

Instructor - Prof. (ret.) Siamak Najarian

Supervisor - Prof. Ilker Hacihaliloglu

Prepared by

Alvin B. Kimbowa

December 9, 2023

Abstract

About 1 in 7 Canadians lives with diagnosed osteoarthritis (OA), with First Nations people having twice the prevalence. Current diagnostic methods, primarily X-ray and MRI, are costly, inaccessible, and pose challenges, particularly in resource-limited settings. Point-of-care ultrasound (POCUS) emerges as a promising solution due to its cost-effectiveness and portability. In this project, I adopted the nnUNet machine-learning algorithm to automatically segment knee cartilage in 2D POCUS devices. I trained the model on knee ultrasound data from over 120 patients. The model achieves an IoU of 0.91, a dice coefficient of 0.95, and a surface dice coefficient of 0.95 at a tolerance of 0.5mm. I then integrated it algorithm into the image processing pipeline of a POCUS device. The model still achieved good qualitative results when evaluated in real-time. This work lays a foundation for numerous machine-learning applications for osteoarthritis management including early OA detection and automated cartilage thickness calculation.

1 Introduction

Osteoarthritis (OA) is a degenerative disease that affects the joints and causes the breakdown of joint cartilage and the underlying bone. It can affect any joint but most commonly affects weight-bearing joints of the spine, hips, knees, and feet. Common symptoms of OA include pain, swelling, stiffness, and loss of flexibility in the affected joint resulting in chronic disability and a reduced quality of life. About 1 in 7 Canadians lives with diagnosed osteoarthritis [1], with First Nations peoples having twice the prevalence [2]. Over $157 billion is spent on OA treatment with an annual productivity loss of over $17 billion expected by 2023 in Canada [3]. This in effect, OA places a significant burden on both the healthcare system and the economy.

To date, there is no cure for OA, and various treatment options aim to manage symptoms and improve joint function. These include medications, physical therapy, lifestyle modifications, and in severe cases, surgical interventions like joint replacement surgery. However, for these therapies to be effective, they need to be started in the early stages of OA when structural damage is not pronounced, and disease progression can be slowed down. They also require continuous monitoring for effective treatment planning. Medical imaging is at the forefront of osteoarthritis diagnosis, treatment, and management as it aids in assessing joint tissues, cartilage, and bone.

Currently, OA is diagnosed based on symptoms and confirmed using X-ray. However, X-ray has poor contrast for soft tissue making early detection of osteoarthritis on radiographs challenging. MRI is thus used to detect subtle changes in the soft tissues such as inflammation or degeneration in the cartilage. Nevertheless, MRI is not suitable for claustrophobic patients or patients with medical implants. Most importantly, however, X-ray and MRI are very costly and inaccessible with long wait times which limits their frequent use, especially in low-resource settings such as the First Nations communities. This results in late diagnosis, undertreatment, and inequality in both clinical care and OA research in such communities.

Point-of-care ultrasound (POCUS) has the potential to mitigate accessibility challenges as it can be used outside of a hospital setting which significantly boosts access to high-risk OA patients. This is owing to the relatively low cost, real-time, and non-ionizing nature of ultrasound. Furthermore, a recent study has shown that ultrasound echo intensity is a valid measure of cartilage composition as t2 relaxation times can be accurately estimated from ultrasound data [4]. This is especially crucial for early detection of osteoarthritis. However, this approach requires segmentation of the cartilage, which is time-consuming, requires a high level of expertise, and exhibits inter- and intra-observer variability. Some semi-automated segmentation approaches have been proposed [5], but these still require some form of manual segmentation. There is thus a need for a fully automated approach to accurately segment joint cartilage in ultrasound images to exploit the approach in [4] and future approaches.

Over the past decade, artificial intelligence (AI), especially machine and deep learning, has shown impressive performance at various segmentation tasks in the medical imaging field. Therefore, in this project, I aimed to develop an accurate machine-learning (ML) algorithm for automated knee cartilage segmentation. Specifically, I aimed 1) to obtain and preprocess knee-ultrasound data, 2) to develop and evaluate a machine learning algorithm for knee-cartilage segmentation, and 3) to integrate the developed ML algorithm in the image processing pipeline of a POCUS device.

Table 1: The dataset used was collected from two institutions, Michigan State University (MSU) and Rutgers University. The MSU dataset consisted of images from patients who had undergone Articular Cruciate Ligament Reconstruction (ACLR) in one knee, images from a healthy control group, and images from healthy volunteering Division 1 Collegiate Female Athletes taken at two visits with two ultrasound devices, the Clarius ultrasound and GE.

Dataset	**Patients**	**No. of images**
ACLR Control	28	84
ACLR	20	119
Rutgers	-	200
Clarius	72	234
GE		246
Total	**120**	**883**

2 Methodology

2.1 Dataset Overview

The dataset used in this project consisted of knee-ultrasound scans of 120 patients from two institutions, Michigan State University and Rutgers. The patients were distributed as in Table 1. For patients who had undergone anterior cruciate ligament reconstruction (ACLR), 6 images were taken: 3 on the ACLR limb and 3 on the contralateral limb. For the rest of the healthy participants, 3 images were taken on the dominant limb. All images were obtained with the knee in the suprapatellar position to expose the knee cartilage using two ultrasound devices. A total of 883 images was obtained and the knee cartilage was manually segmented by a single expert reader. I preprocessed the images to remove unwanted metadata as shown in Fig. 1. I then randomly split the data into two sets, one to be used for training the machine learning model, and the other for evaluating the model performance. The training set consisted of images from 80% of the patients (96 patients) and the test set consisted of images from the remaining 20% of the patients (24

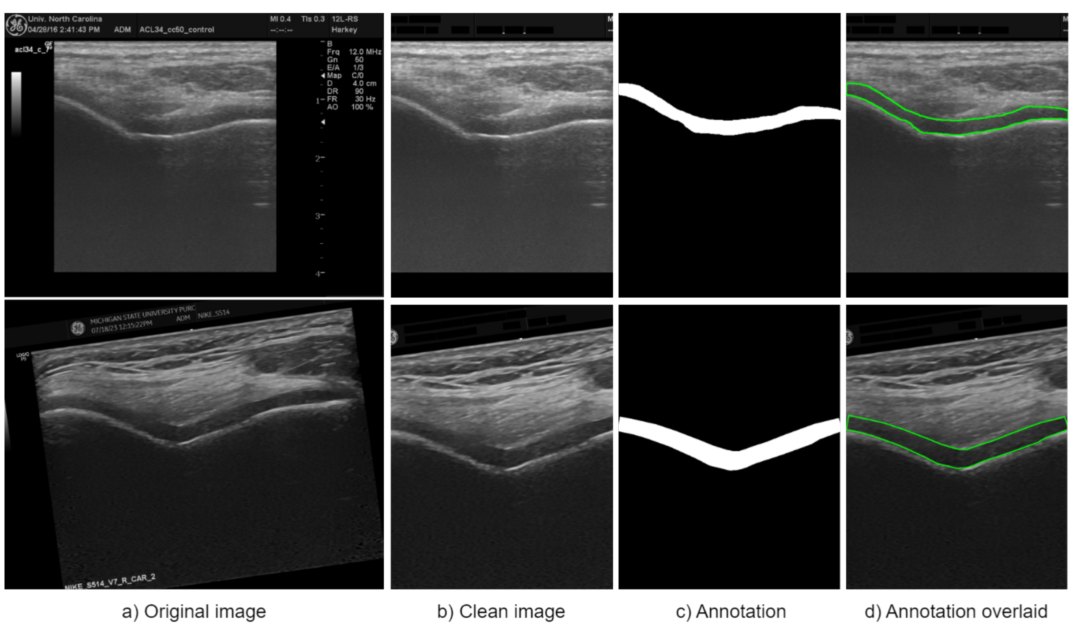


Figure 1: The dataset used consisted of B-mode ultrasound images together with expert annotations of the knee cartilage. The original images a), had metadata that was removed using a combination of inpainting and cropping to yield clean images b). The cartilage segmentations c) were manually made by a single expert reader. In d), the annotation is overlaid on the clean image.

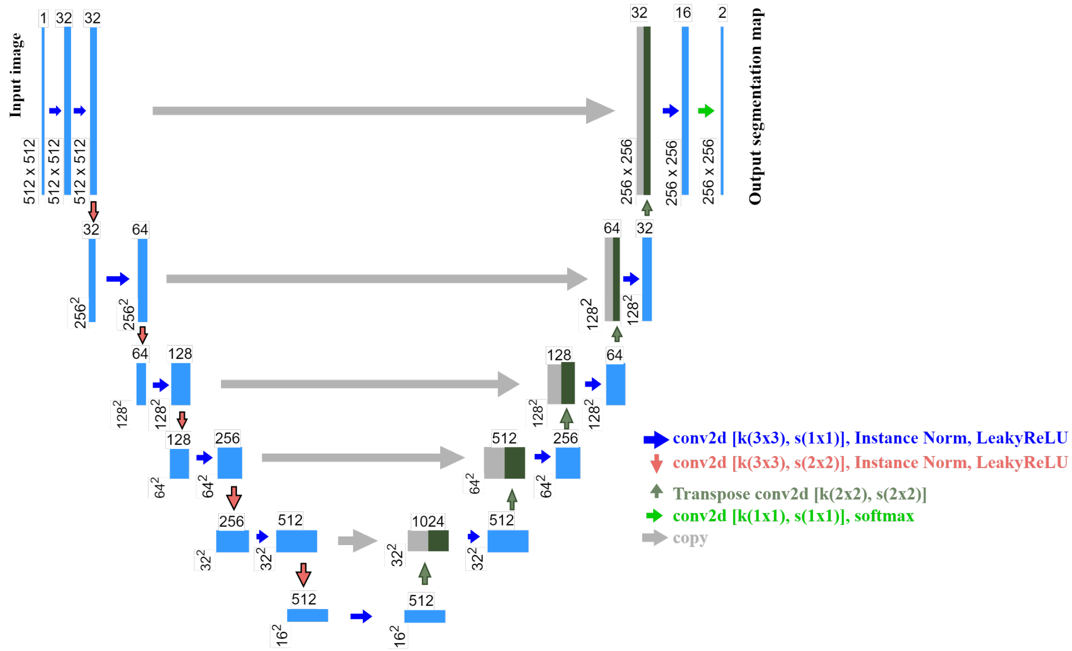


Figure 2: A visual representation of the best nnUNet architecture adopted in this project. The model takes as input a 2D B-mode ultrasound image and outputs a 2D segmentation map of knee cartilage. The model has no pooling layers as is the case in the original UNet, but rather uses convolutions with a stride of 2 to half the image size in the encoder.

patients).

2.2 Model development

I adopted the nnUNet model architecture, a state-of-the-art machine-learning algorithm for medical image segmentation [6]. Fig. 2 shows a visual representation of the nnUNet model architecture. The model takes as input a 2D B-mode ultrasound image, encodes the information within the image via the downward path, and decodes the information via the upward path generating a segmentation map highlighting the knee cartilage. The model was compiled with a compound loss function which was a weighted sum of the binary cross entropy loss and dice loss. The optimizing metric used during training was the dice coefficient, that is, the best model was one that yielded the highest dice coefficient on the validation data. The algorithm was implemented using the open-source code provided in [6].

I trained the model using 705 images from the 96 patients. During training, I performed a 5-fold cross-validation whereby I further split the training data into 5 equal folds, and at a given time, I trained the model on 4 fours and evaluated it on the remaining fold. I did this 5 times (splits) such that each fold was used for testing as shown in Fig. 3. For each split, the model was trained for 1000 epochs on the NVIDIA Tesla V100 8GB NVLink GPU provided by UBC's Advanced Research Computing Sockeye. Finally, after obtaining the average mode performance across all splits, I retrained the model on all five folds. This approach provides a more robust evaluation of the model performance by removing bias that could arise as a result of the data split used. It is also an efficient use of data as compared to leaving out a validation set, in addition to the test data, that would not be used for training the model.

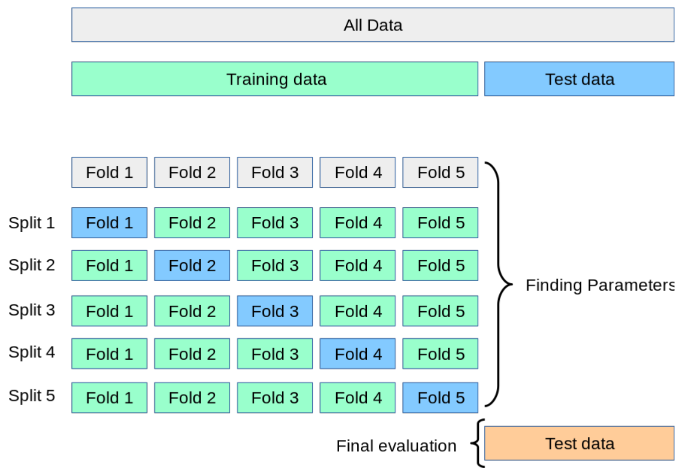


Figure 3: The dataset was split into two, 80% of the patients' data was used for training the model, and 20% was used for model evaluation. The training data was further split into five equal stratified folds to perform a 5-fold cross-validation. For each split, the model was trained on four folds and evaluated on 1 fold.

3 Results

3.1 Qaulitative and Quantitative

Fig. 4 shows a sample of the model results for 6 ultrasound images from the test set. For each row, the first image is the original B-mode image followed by the model prediction. The green contours highlight the ground truth annotations, and the red contours highlight the model prediction. It can be seen that the model predictions closely follow the ground truth for Fig. 4a) and Fig. 4b). It was also observed that in some cases, as in Fig. 4c), the model predictions are less accurate.

To quantify the model performance on the test data, I computed various metrics as shown in Table 2 and visualized in Fig. 5. The intersection over union (IoU) penalizes under- and over-

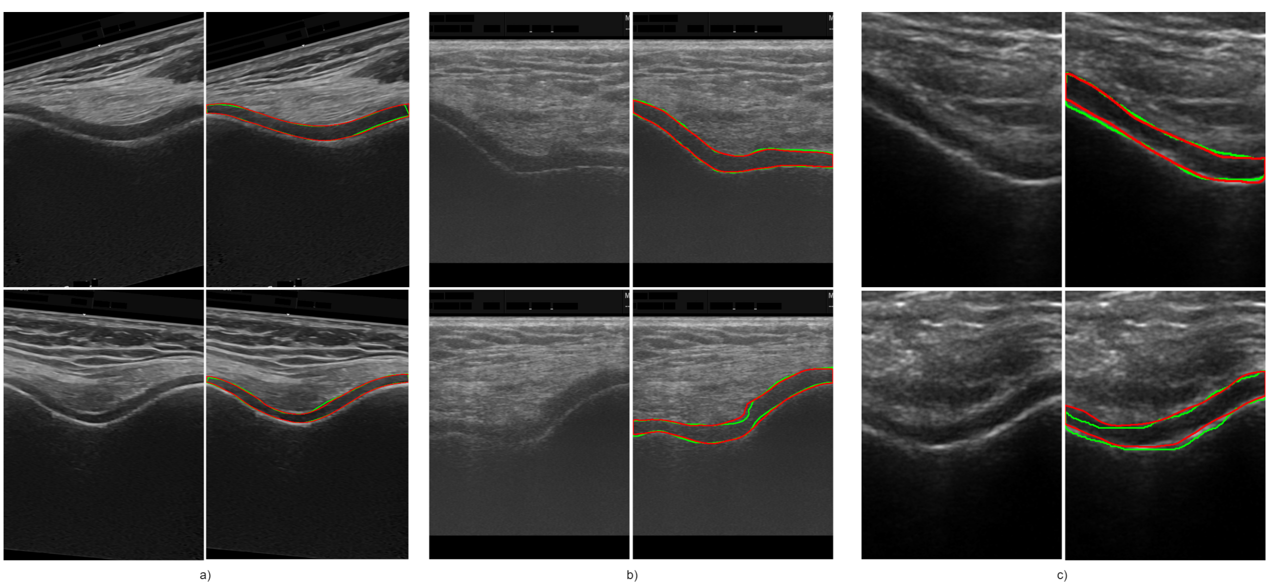


Figure 4: Sample results on unseen test data, where the green contours are the ground truth annotations and the red contours are the model predictions. The model predictions closely align with the ground truth regardless of the tilt direction in the image a) or the shape of the cartilage b). However, larger errors are observed in some cases c).

Table 2: Evaluation metrics of the nnUNet model across all the datasets. The mode exhibits robust performance as can be seen by the overall surface dice-Sorensen coefficient of over 0.95 within a tolerance of 0.5mm, and a Huasdorff Distance of 0.52mm at a 95^{th} percentile.

Dataset	IoU	Dice	Surface Dice		Hausdorff Distance	
			@0.5mm	@1mm	@95%	@100%
ACL	0.92	0.96	0.99	1.00	0.22	0.44
Control	0.91	0.95	1.00	1.00	0.17	0.31
Clarius - auto	0.90	0.95	0.88	0.96	1.42	4.24
Clarius - std	0.91	0.95	0.87	0.97	0.87	2.77
GE - auto	0.93	0.97	0.98	0.99	0.25	1.26
GE - std	0.93	0.96	0.99	1.00	0.27	1.15
Rutgers	0.84	0.91	0.95	1.00	0.43	0.59
Overall	**0.91** ±0.03	**0.95** ±0.02	**0.95** ±0.06	**0.99** ±0.02	**0.52** ±0.46	**1.54** ±1.45

segmentation. The dice coefficient, also known as the F1 score, penalizes false positives. IoU and dice are important when evaluating how much of the cartilage was predicted to be cartilage by the model. The surface dice and Hausdorff distance, on the other hand, measure the overlap between the boundaries of the model prediction and the ground truth. They are especially important in knee-cartilage segmentation as it is important to accurately capture the boundaries of the cartilage for downstream tasks, such as t2 relaxation time estimation, to be successful.

The model achieves high IoU and Dice scores, above 0.9, across all the datasets which is acceptably great performance. This can be attributed to the fact that the knee cartilage is conspicuous within the B-mode ultrasound image. The model also achieves a very high surface dice score of 0.95. This signifies that the predicted cartilage contours closely align with the ground truth within a tolerance of half a millimeter.

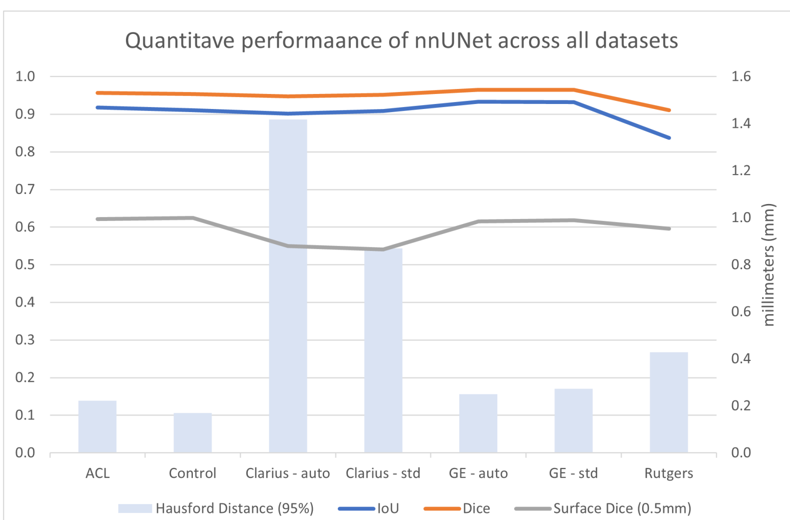
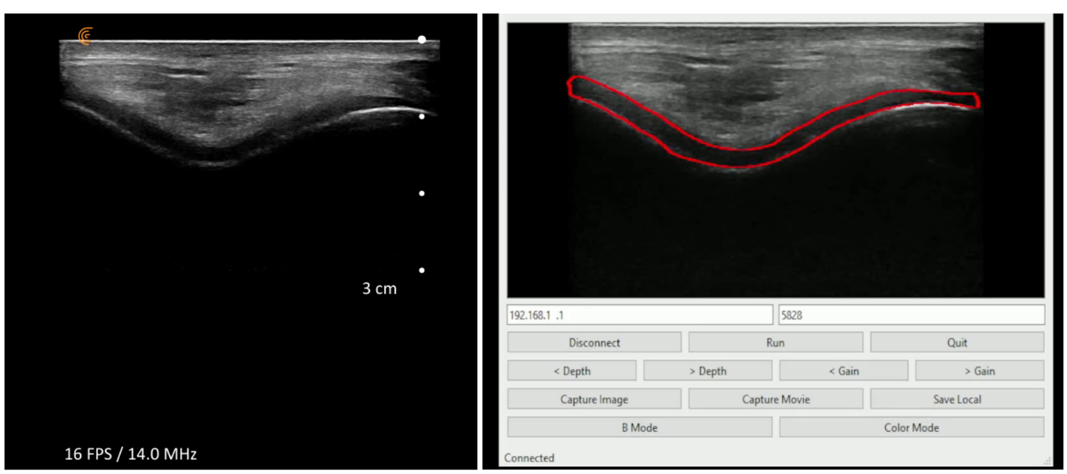


Figure 5: The Haudorff Distance was the largest for the Clarius datasets and this could be due to a data distribution shift across various scanners and the model is not robust enough to deal with these variations.

Figure 6: The Clarius L15 HD3 probe used in the project. The probe can cast images to Android and PC wirelessly. Clarius also provides an API that can be used to integrate machine learning algorithms within the image processing pipeline of the device.

3.2 POCUS Integration

I integrated the model into the image processing pipeline of the Clarius L15 HD3 POCUS device shown in Fig. 6. The Clarius ultrasound combines the transducers and all the necessary image-processing components within a handheld probe. Images from the probe can be streamed wirelessly within the Clarius mobile application on a smartphone or on a laptop. Fig. 7 is a snapshot from a real-time demo of me taking a knee ultrasound scan with the knee in the suprapatellar position. Without the machine learning algorithm, a sonographer would see B-mode images as in Fig. 7a). Fig. 7b) shows a graphical user interface on a laptop computer with the machine learning model integrated within the image processing pipeline to generate the red contour highlighting the knee cartilage in real-time. It can clearly be seen that the model achieves qualitatively good performance. However, the model inference speed on CPU of 0.4 frames per second (fps) is orders of magnitude lower than the operating frame rate of the Clarius which is 16fps. There is thus need to improve inference time for real-time performance



a) Original video as seen in the Clarius android application

b) Model integration as seen on a graphical user interface on a laptop

Figure 7: A snapshot from a real-time integration of the machine learning algorithm in the image processing pipeline of the Clarius HD L15 ultrasound. This was taken from a demo knee-ultrasound in the suprapatellar position. It can be seen that the model still achieves impressive qualitative performance.

4 Conclusions

Osteoarthritis management requires early detection and continuous monitoring and thus requires an affordable but effective imaging solution. Point of care ultrasound (POCUS) has to potential to address this challenge and by leveraging AI, POCUS can find great utility, especially in low-resource settings. Joint cartilage segmentation is at the basis of numerous AI applications for OA management, from automated cartilage thickness measurement to early OA detection. In this project, I developed, evaluated, and deployed a machine-learning algorithm for automated knee cartilage segmentation. The model achieved good segmentation results with an IoU of 0.91, a dice coefficient of 0.95, and a surface dice coefficient of 0.95 at a tolerance of 0.5mm. The model achieved impressive qualitative performance when integrated into the image processing pipeline of the Clarius HD3 L15 device.

Nevertheless, the inference time of the model on CPU was very low, about 0.4 frames per second, as compared to the operating frame rate of the Clarius device, of 16fps. Future work will investigate and or develop lightweight algorithms for real-time inference in POCUS devices given their low computing capability. It will also be interesting to apply the approach in [4] on the segmented knee cartilage. Furthermore, using the segmented knee cartilage, further deep-learning algorithms could be developed to generate OA biomarkers from B-mode ultrasound images.

Bibliography

[1] P. H. A. o. Canada, “Osteoarthritis in canada.” Last Modified: 2020-12-24.

[2] C. Barnabe, B. Hemmelgarn, C. A. Jones, C. A. Peschken, D. Voaklander, L. Joseph, S. Bernatsky, J. M. Esdaile, and D. A. Marshall, “Imbalance of prevalence and specialty care for osteoarthritis for first nations people in alberta, canada,” vol. 42, no. 2, pp. 323–328. Publisher: The Journal of Rheumatology Section: Article.

[3] J. J. Young, O. Važić, and A. C. Cregg, “Management of knee and hip osteoarthritis: an opportunity for the canadian chiropractic profession,” vol. 65, no. 1, pp. 6–13.

[4] M. S. Harkey, N. Michel, C. Grozier, J. M. Slade, K. Collins, B. Pietrosimone, D. Lalush, C. Lisee, I. Hacihaliloglu, and R. Fajardo, “Femoral cartilage ultrasound echo-intensity is a valid measure of cartilage composition,” vol. n/a. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jor.25722.

[5] M. S. Harkey, N. Michel, C. Kuenze, R. Fajardo, M. Salzler, J. B. Driban, and I. Hacihaliloglu, “Validating a semi-automated technique for segmenting femoral articular cartilage on ultrasound images,” vol. 13, no. 2, p. 19476035221093069. Publisher: SAGE Publications Inc.

[6] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-net: a self-configuring method for deep learning-based biomedical image segmentation,” vol. 18, no. 2, pp. 203–211. Number: 2 Publisher: Nature Publishing Group.